

Neural Mechanisms for Heading and Structure-from-Motion Perception

R.A. ANDERSEN, D.C. BRADLEY, AND K.V. SHENOY

*Division of Biology, California Institute of Technology,
Pasadena, California 91125*

Two of the most important perceptual functions of the visual motion system are to compute our direction of heading as we move through the environment, and to deduce the three-dimensional structure of objects and the environment from motion cues. Below, we review experiments that provide insights into how these perceptual phenomena are constructed by the brain. Understanding how the motion system performs these analyses will likely have general applicability to other perceptual functions, both within and outside the motion pathway. For instance, understanding how motion signals are perceived as spatially constant despite eye movements, an important prerequisite for determining heading direction, may lead to a general understanding of spatial-perceptual constancy. Likewise, understanding how three-dimensional form is processed from motion cues in the dorsal visual pathway may provide important suggestions as to how form is derived from other visual cues in the ventral visual pathway.

In the first part of this paper, we discuss how "optic flow" signals, which are generated by translation through the environment, can be used for navigation. In particular, an important question in this area is how flow generated by eye or head movements is subtracted from the flow generated by translation in order to recover the direction of heading. Current studies indicate that the dorso-medial superior temporal area (MSTd) plays a pivotal role in heading computation.

In the second part of the paper, we examine the neural basis of structure-from-motion (SFM) perception. In particular, evidence is reviewed indicating that the middle temporal area (MT) is a site for the extraction of three-dimensional structure from motion cues.

HEADING COMPUTATION BY AREA MSTd

Theories of Heading Computation

Gibson (1950) proposed that observers in motion can use the focus of expansion of the retinal image to determine the direction of heading. If the eyes and head do not move, then the focus of expansion corresponds to the direction of heading, and navigation can be achieved by maintaining that focus in the desired direction of locomotion. Regan and colleagues and other investigators have pointed out that such a

solution only works in the special case of no eye or head rotations (Longuet-Higgins and Prazdny 1980; Regan and Beverley 1982; Koenderink and Van Doorn 1986; Royden et al. 1992; Warren 1995). This point is made in Figure 1 for the simple case of moving toward a wall. If the eyes are moving, such as would occur while fixating a feature on the wall that is displaced from the heading direction, then a roughly laminar motion (which is opposite in direction to the eye movement) is added to the expansion component. As a result, the retinal expansion focus is shifted in the direction of eye movement. The true focus of expansion could be recovered by decomposing the flow field into its components due to translation (forward movement) and rotation (eye movement). After decomposition, the focus of the expansion component would indicate the direction of heading (Longuet-Higgins and Prazdny 1980). The same principles apply for scenes with depth variation, although the flow pattern during eye movements can be more complex.

Two general methods can be used to accomplish the decomposition of the translational and rotational components of flow. One uses information from the retinal image, such as local speed differences (motion parallax), and the other uses extra-retinal cues, such as a signal related to pursuit eye movements. There have been several different computational proposals to explain how this decomposition is performed (Longuet-Higgins and Prazdny 1980; Koenderink and Van Doorn 1981; Rieger and Lawton 1985; Hildreth 1992), and all of these models use retinal cues in the flow field to perform this decomposition. In general, these computational models use depth variation in the display to separate the expansion and laminar components of flow. This decomposition is made possible by the fact that translation through the world causes differential retinal motions as a function of depth, whereas rotation of the eyes or head causes laminar motion which is the same across the visual field and does not depend on depth. Therefore, if the image contains depth differences, it is possible to tease apart the flow components due to translation and rotation.

The above, earlier models developed rather general algorithms for solving the optic flow problem, and as such, were generally not specific about how they would be implemented by the brain. Subsequently, several models have been developed to perform the decom-

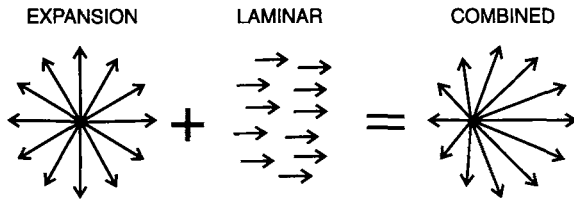


Figure 1. The problem of navigating while our eyes are moving. Moving toward a flat surface, we see an expanding image (*left panel*), and as long as our eyes are still, the focus of this expansion corresponds to our direction of heading. However, if we move our eyes to the left—as we would, e.g., while tracking an object off to the side—this adds rightward laminar flow to the retinal image (*middle panel*). The combined retinal image (*right panel*) is similar to the original expansion but has its focus shifted. Therefore, when the angle of gaze is changing, the retinal focus no longer corresponds to the heading. (Reprinted, with permission, from Bradley et al. 1996a [copyright AAAS].)

position of eye rotation and observer translation components in a biologically realistic manner (Hatsopoulos and Warren 1991; Lappe and Rauschecker 1994; Perrone and Stone 1994; Warren 1995; Stone and Perrone 1996). These are generally referred to as template models because they use receptive fields as templates to match various flow patterns. These models contain many maps (sets of templates) for the many different optic flow conditions that can arise from different eye rotation speeds, translation speeds, and depth variations and structure in the scene. A criticism of this template approach is that it requires too many maps (and neurons) to be biologically plausible. Our findings indicate that optic-flow-sensitive neurons adjust their focus tuning to account for tracking eye movements, which greatly reduces the number of templates required and makes the general template approach more realistic. This adjustment is made using an extra-retinal eye-movement signal, the method for computing heading which has not received much attention in computational models.

Psychophysics

For a long time after Gibson's original proposal, it was believed, on the basis of psychophysical studies, that humans were not very accurate at using flow information for navigation. This issue was later re-examined by Warren and colleagues using more controlled psychophysical techniques. When their stimuli imitated the approach to a wall (i.e., no relative depth cues), they found that heading judgments were very accurate under conditions of no eye or head rotations (for review, see Warren 1995). These judgments remained accurate during pursuit eye movements, even though these add laminar flow to the stimulus and shift the focus away from the heading. However, if the eye movements were simulated by adding laminar flow directly to the stimulus, subjects were unable to com-

pensate for the laminar flow and thus could not accurately judge their heading. These results, later confirmed by Banks and colleagues (Royden et al. 1992, 1994), suggested that an extra-retinal (pursuit) signal is needed to decompose the retinal image into its radial and laminar components.

The psychophysical studies discussed above were based on the simulated approach to a wall. Different results were obtained when stimuli contained motion parallax cues—differential motions created by depth differences in the stimulus. In some cases, these parallax cues allowed observers to correctly judge their heading movements (Warren and Hannon 1988; Warren 1995), whereas in other conditions they could not (Royden et al. 1992, 1994).

Physiology of Area MSTd: Optic Flow and Smooth Pursuit Sensitivity

Area MSTd cells have been found to respond to rotations, expansion/contractions, and laminar motion (Sakata et al. 1985; Saito et al. 1986; Tanaka et al. 1986, 1989; Tanaka and Saito 1989; Duffy and Wurtz 1991a,b; Graziano et al. 1994; Lagae et al. 1994). The observation that many MSTd cells are selective for these types of stimuli led many observers to propose that area MSTd is involved in navigation from optical flow analysis. Moreover, the receptive fields of the expansion-selective neurons are also tuned for the location of the expansion focus, providing additional support for the idea that MSTd plays a role in navigation from optic flow (Duffy and Wurtz 1995). Another important aspect of MSTd neurons is their invariance for location, scale, or form cues (Graziano et al. 1994; Geesaman and Andersen 1996). In other words, a cell preferring an expansion will be tuned best to an expansion stimulus (e.g., as opposed to a rotation stimulus) regardless of its location in the receptive field, its size, or the environmental features conveying the motion signal. This high degree of invariance is important for accomplishing navigation independent of the exact structure of the visual environment and provides strong support for template models.

Several studies have documented activity in the medial superior temporal area (MST) related to smooth pursuit eye movements (Mountcastle et al. 1975; Lynch et al. 1977; Komatsu and Wurtz 1988a,b; Newsome et al. 1988). This coincidence of optic flow selectivity and pursuit selectivity in MSTd suggests that this area is a prime candidate for a brain center involved in heading computation. Area MST also receives angular-rotation vestibular signals, which are generated during head pursuit (Kawano et al. 1980, 1984; Kawano and Sasaki 1984; Thier and Ericksen 1992). The vestibular component during head pursuit has been measured with vestibulo-ocular reflex (VOR) cancellation, a condition in which the monkey maintains fixation on a stimulus that is attached to a rotating vestibular chair in which he is seated. The preferred

direction of activity during VOR cancellation typically corresponds to the preferred eye pursuit direction, suggesting that the cells are coding gaze-tracking movements, independent of whether eye or head rotations lead to the gaze pursuit. Thus, area MSTd may also provide heading compensation during head, as well as eye, movements.

Heading Computation during Eye Movements

Recently, we examined the simplest case of heading computation from a physiological perspective, translation of an observer toward a wall (i.e., a frontoparallel plane) (Bradley et al. 1996a,b). Compensation for pursuit in this condition has been shown to occur for human observers when subjects are pursuing, but not when pursuit is simulated with an identical retinal stimulus while the eye is held fixed. We examined the focus tuning of area MSTd neurons with a stimulus on the screen corresponding to observer translation toward a wall; i.e., expansion with no depth variation. If MSTd is involved in heading computation, we would expect focus-tuning curves to shift during pursuit eye movements in order to continue to code the correct heading direction with the eyes moving. We would also expect from the psychophysical results that movement compensation would not be present when the same retinal stimulus is generated as the pursuit condition, but without eye movements. We found that both of these predictions are true, providing strong evidence that area MSTd plays a direct role in heading computation.

In these experiments, monkeys were seated with heads fixed and trained to fixate a spot of light on a projection screen. The spot was stationary or moving at 15.7°/sec in a constant direction. When the spot moved, the monkey was required to pursue it. Neurons were stimulated, during pursuit or stationary fixation, with a visual expansion pattern made up of 600 moving dots. The expansion stimulus was 130° × 130° square, but only a 50° × 50° segment (window) of this larger pattern was shown in a given trial (see Fig. 2).

After isolating an MSTd neuron and mapping its receptive field, we determined the neuron's preferred optic flow pattern, i.e., expansion, contraction, clockwise rotation, or counterclockwise rotation, by showing these patterns in pseudorandom order and noting the strongest response. If the cell responded to expansion, we proceeded with the experiments below. If it responded to one of the other patterns, we proceeded with the experiments outlined in a subsequent section. We then determined the neuron's preferred pursuit direction by requiring the monkey to pursue a dot across the screen in eight different directions: 0°, 45°, ..., 315°. The direction eliciting the strongest response was taken as the preferred pursuit direction, with the opposite direction being designated the "antipreferred" pursuit direction. Next, nine expanding patterns were shown, in each case shifting the expansion focus (origin) to a new location. The focus varied from

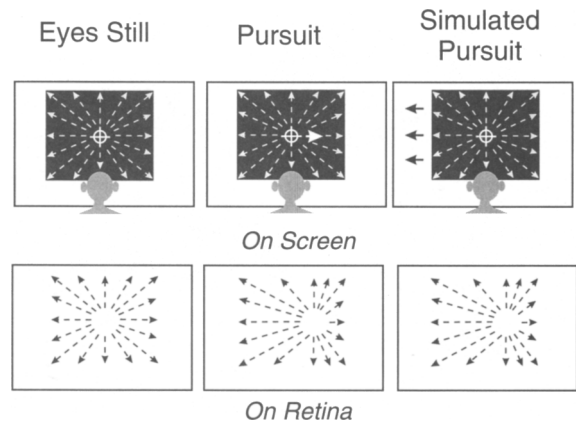


Figure 2. Paradigm to test whether MSTd cells encode the heading. A series of expanding images are shown while the monkey fixates (*top left*), during which time we record single-unit MSTd activity. The expanding patterns differ in terms of their focus position, allowing us to measure each neuron's output as a function of focus position. The focus position is the same on the screen and on the retina with eyes still (*bottom left*). The same series of expansions is also shown while the monkey pursues a moving target over a short distance (*top middle*). The eye position is about the same as in the fixed-eye case, but the eye velocity displaces the retinal focus relative to the screen focus (*bottom middle*). If a given neuron encodes the heading, its activity should be the same for a given screen focus position, regardless of whether the eyes are moving or still. In control experiments, eye movement is simulated by holding the eye still while moving the stimulus (*top right*). Like the pursuit condition (*middle panels*), this shifts the retinal focus (*bottom right*), but now no eye movement signal is available.

−40° to 40° in 10° increments, along an axis parallel to the neuron's preferred-antipreferred pursuit axis. Focus shifts were accomplished by shifting the 130° × 130° pattern behind the 50° × 50° viewable window, which remained at the same screen location. Thus, the stimulus always appeared in the same location on the screen (approximately centered in the neuron's receptive field); only the position of the focus, relative to this window, varied.

While the above stimuli were being shown, the monkey was required to fixate a stationary spot or to pursue a spot moving in the preferred or antipreferred direction for the cell (Fig. 2, left and center panels). Data were collected for the middle 0.5 sec of the stimulus presentation interval (1 sec), and the starting point of the moving spot was set so that the average position of the spot during the data collection interval was the same as the spot's position during stationary fixation. Therefore, the position of the eyes was about the same in both conditions (fixation and pursuit), differing at most by 4° during the data collection period. However, during pursuit, the retinal focus position was shifted relative to the focus position on the screen (Fig. 2, compare lower left and lower center panels). The magnitude of this shift was 30° for all stimuli in the

direction of pursuit, since at this eccentricity, dot speeds equal the pursuit speed (15.7°/sec; see above). In other words, a dot moving at the pursuit speed (and in the same direction) was stationary on the retina and thus became the new focus.

For each neuron, we defined three focus-tuning curves: one for stationary fixation, one for pursuit in the neuron's preferred direction, and one for pursuit in the neuron's antipreferred direction. Each curve expressed the neuron's firing rate as a function of the position of the focus. Since each focus position can be expressed in two ways—the position of the focus on the screen, or the position on the retina—our basic task was to find the coordinate frame (screen or retina) in which the three tuning curves are best aligned.

To illustrate, Figure 3 shows the basic result of the experiment. The upper panels show focus-tuning

curves for fixation and preferred/antipreferred pursuit eye movements. The left panel expresses these curves in screen coordinates (recall that the screen focus is equivalent to heading), and the right panel expresses the curves in retinal coordinates. All three curves are well aligned in screen coordinates but are shifted relative to each other in retinal coordinates. In other words, this neuron tends to give the same response to a given screen focus position, regardless of whether or not the eyes are moving. In contrast, this neuron's response to a given retinal focus varies considerably, depending on the state of the eyes. Therefore, neurons such as this could compute heading by shifting their retinal tuning curves in such a way as to maintain a constant output for a given screen focus (direction of heading).

Shifts in retinal tuning curves could be caused either

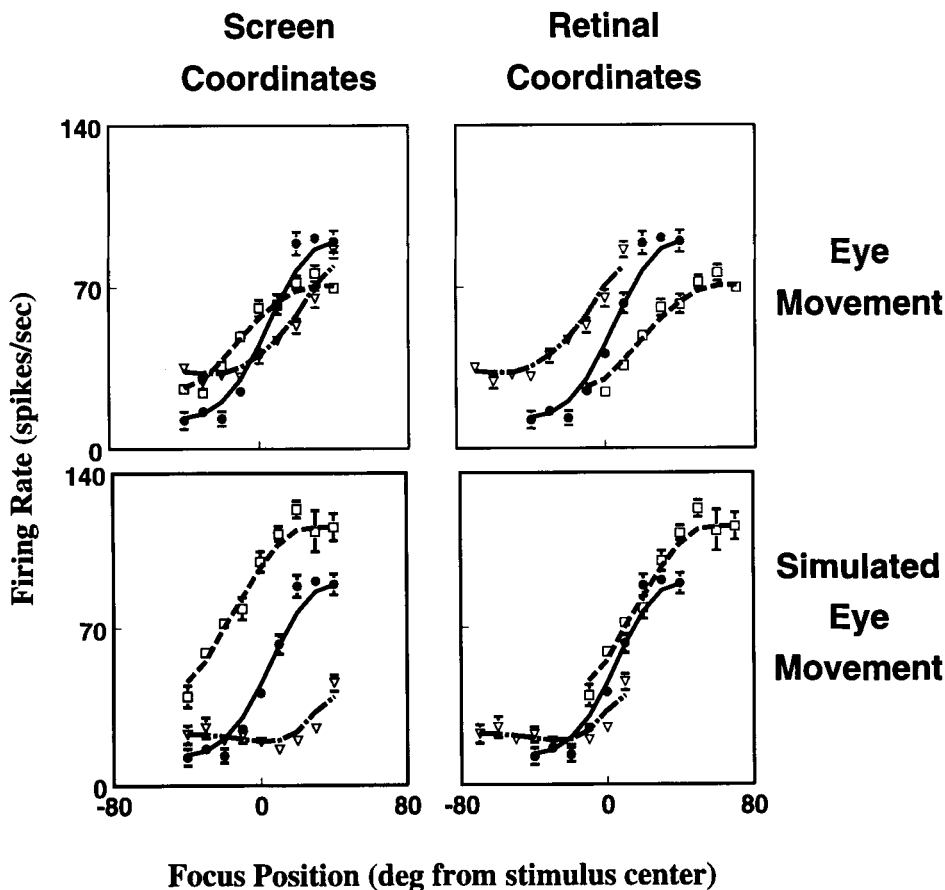


Figure 3. Example of an MSTd "heading cell." In all panels, the solid lines/solid circles represent the fixed-eye focus tuning (identical in all four graphs), dashed lines/open squares are preferred-direction eye movements (real or simulated), and dot-dashed lines/open triangles are antipreferred-direction eye movements (real or simulated). (*Top row*) Pursuit eye movements. (*Bottom row*) Simulated pursuit eye movements. (*Left column*) Screen coordinates. (*Right column*) Retinal coordinates. Data in the left and right columns are the same, except pursuit curves in the right column were shifted by $\pm 30^\circ$ relative to screen coordinates to give retinal coordinates. When responses are expressed in screen coordinates (*top left*), activity is roughly constant for a given focus position. Since the screen focus corresponds to the heading, this implies that neurons such as this could encode heading direction. Note that for simulated eye movements, results are quite different (*bottom right*), so an extraretinal signal is essential for computing the heading. Data points are means \pm S.E.M. for four replicates, where each replicate is the mean firing during the middle 500 msec of the stimulus-presentation interval. (Reprinted, with permission, from Bradley et al. 1996a [copyright AAAS].)

by eye movements directly, or by what the retina sees during that movement. To distinguish these possibilities, we used a simulated eye movement control. For example, to simulate smooth pursuit to the right, we moved the entire stimulus left while holding the eyes still (Fig. 2, upper right panel). Thus, the retinal image was identical for real and simulated pursuit, the only difference being the actual velocity of the eyes (Fig. 2, compare lower center and lower right panels). The lower panels in Figure 3 show simulated eye movement data for the neuron described immediately above. In this case, tuning curves are well aligned in retinal coordinates but highly dissimilar in screen coordinates. Thus, the shift in retinal tuning exhibited by this neuron was due to the eye movement itself, not the retinal consequences of the movement.

To quantify results in the MSTd population as a whole, we used cross-correlation techniques to find the shift in retinal focus tuning caused by pursuit. In other words, focus-tuning curves for pursuit and stationary fixation were incrementally shifted relative to each other (in retinal coordinates), in each case correlating the two. The relative shift at which maximal cross-correlation occurs was taken as the "best" shift—that which best aligns relevant features of the two curves, such as maxima, minima, and inflections. These best shifts were then summarized in a frequency histogram for the population.

We have cross-correlation data for 57 expansion-selective MSTd neurons. When eye movements were simulated, the frequency distribution clearly peaks at zero, indicating that in the absence of real eye movements, retinal focus-tuning curves generally do not shift. On the other hand, the histogram for real eye movements was clearly spread out to one side, so that more than half of the neurons have focus-tuning curves shifted by 10° or more, in every case in the direction required to compensate for the retinal focus shift induced by the eye movement (average shift = $11 \pm 3^\circ$ relative to simulated pursuit; $18 \pm 3^\circ$ when shifts are computed based on mean square difference between the curves). Our findings therefore indicate that MSTd cells integrate eye movement signals with optic flow in order to shift their retinal focus sensitivity, thus compensating for retinal focus shifts incurred during eye movements.

These results indicate that many MSTd neurons shift their focus tuning by the right amount, while others do not at all; a third group is intermediate. There are two possible interpretations of these results: (1) Only a portion of the MSTd neurons are involved in heading computation (those with complete compensation), whereas the others are involved in some other function (these being the ones that do not compensate), or (2) the distribution of shifts reflects the transformation from retinal to heading coordinates. In the latter case, we would expect to find cells in different stages of the transformation, i.e., early (non-shifted), intermediate (partly shifted), and final

(fully shifted), which is indeed what we found. The following section suggests a mechanism for this transformation.

Gain Effects and a Model for Pursuit Compensation

As mentioned above, results indicate that a fraction of cells in MSTd do not have focus-tuning shifts with pursuit. However, most of these cells do show modulation of their amplitude tuning by pursuit, by an average of $25 \pm 3\%$ for the cells in our sample. This finding suggests a possible model of how the shifting might be accomplished using an intermediate step of gain-modulated neurons without shifting fields (Bradley et al. 1996a).

The model for the transformation from retinal to screen coordinates uses two first-stage neurons whose focus tunings are offset, and whose activity is summed by a second-stage neuron that shifts its focus tuning during pursuit (see Fig. 4). The first-stage cells have nonshifting focus-tuning fields which are sine functions. Sine functions are used because different parts of a sine function can approximate either a gaussian or a sigmoid, which are the general shapes of the MSTd focus-tuning curves. Each sine function has three parameters: amplitude, frequency, and phase, as well as two gain parameters that are applied during each of the opposed pursuits. All parameters were adjusted simultaneously, fitting the fixed-eye and two pursuit directions concurrently, using nonlinear least-squares regression. This analysis was performed on the data from all cells that show focus-tuning shifts. This analysis accounted not only for shifts in focus-tuning curves, but also for gain modulation of the magnitude of activity for the shifting neurons. We have performed this analysis on 36 neurons that clearly shift their focus tuning during pursuit. The fits have been extremely good for such a simple model ($r^2 = 0.72 \pm 0.03$) and much better than single-stage models that use nonlinear pursuit modulations of the focus-tuning curves (sigmoids, exponentials, thresholded linear functions).

The above model demonstrates how MSTd can correct for the motions due to eye rotation. However, neither this model nor the experimental results can at this point determine in which coordinate frame MSTd represents the heading. In the above study, the position of the eyes, head, and body were all aligned and we referred to the MSTd response as being converted to "screen coordinates" only as an operational term. Additional experiments will be needed to dissociate these different coordinate frames by performing the experiments with the eyes, head, and body in different positions with respect to one another. One would presume that at some point in the nervous system, heading direction is represented in body and world coordinates in order to permit one to walk or drive through the world.

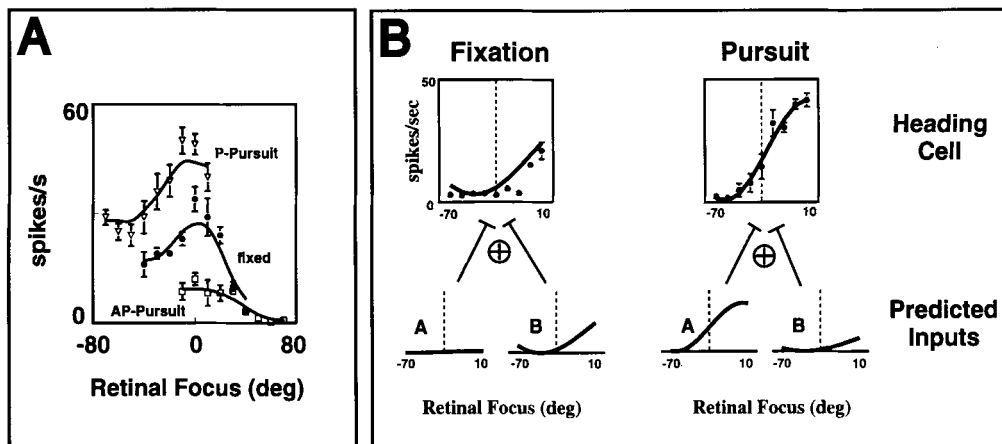


Figure 4. (A) Example of a neuron whose focus tuning does not shift during pursuit, but instead changes in amplitude. The three curves, corresponding to stationary fixation and opposite directions of pursuit, all peak near 0° , but their amplitudes vary considerably. Cells like this could serve as input to “heading cells,” whose focus tuning does shift during pursuit. (B) Example of a heading cell and how it might be constructed from nonshifting, gain-modulated cells. The upper graphs show the neuron’s measured focus tuning, during fixation (left) and preferred-direction pursuit (right). (Circles) Mean response; (curves) model fit. (Lower panels) Predicted input functions. Each function is characterized by three sine-wave parameters and multiplied by a gain; the two functions are summed to make the focus-tuning curve of a heading cell. All parameters were adjusted by nonlinear regression to fit the data (upper panels). This neuron’s focus tuning shifts leftward during pursuit, and the model recreates this shift by increasing the gain on input function A while decreasing the gain on input function B. The sine-wave parameters (other than gain) were identical for the fixation and pursuit conditions; only the gains were adjusted to simulate the focus-tuning shift. (Reprinted, with permission, from Bradley et al. 1996a [copyright AAAS].)

Heading Computation during Head Movements

The results from the previous section indicate that an eye velocity signal is needed to compensate for smooth pursuit eye movements during heading judgments. A head velocity signal would also enable heading direction to be calculated during head movements. In conjunction, the eye and head velocity signals would permit the computation of heading from optic flow during arbitrary eye and head motions, especially in situations where a part of the gaze pursuit is supported by head movement and part by eye movement. It is presently not known if MSTd neurons compensate during head movements, or even if human subjects can perceive heading direction during pursuit head movements. If head pursuit compensation is present, there are at least three potential sources for the head movement signal: efference copy, the vestibular system, and neck proprioception. As mentioned above, vestibular as well as eye pursuit signals have been recorded in area MSTd. In preliminary experiments in which a monkey is rotated in a vestibular chair, we have found evidence that at least some cells show focus-tuning compensation during head movement as well as eye movement (Shenoy et al. 1996). Since the animals were rotated with fixed head-on-body orientation in the chair, the source of the compensation originates from vestibular canal signals. This result suggests that MSTd may also play a role in heading computation during head movements.

Rotation, Contraction, and Spiral Patterns

Area MSTd contains cells sensitive not only to expansion, but also to rotation, contraction, laminar flow, and spiral motions. During self-motion, these various motion patterns can be generated depending on the structure of the environment and the movement of the eyes. For instance, when fixating a location forward and to the side on the ground plane while translating forward, the motion around the fixation point is in the form of a spiral (Graziano et al. 1994; Warren 1995). Such nonexpansion stimuli can potentially provide information for computing heading. However, non-expansion stimuli suffer similar problems when the eyes are moving; i.e., their focus shifts on the retina during pursuit movements. Interestingly, these shifts are not in the direction of the eye movement as in the case of expansion. For rotation, the focus shifts orthogonally to the direction of pursuit (see Fig. 5), the contraction focus moves opposite to the direction of pursuit, and spiral foci move in oblique paths that depend on the rotation and expansion/contraction components producing the spiral.

In the previous experiments, we have only discussed expansion stimuli. However, we have also investigated the effects of pursuit eye movements on 36 rotation and 46 contraction neurons. The speed of the dots was the same as the expansion stimuli, although each dot motion was oriented at 90° (rotation) or 180° (contraction) from the expansion case. These cells showed simi-

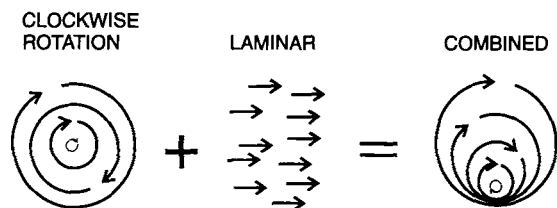


Figure 5. Eye movements across rotating patterns cause an orthogonal shift in the rotation center. For example, the focus in a clockwise rotation shifts downward during a leftward eye movement. To correct for this shift, clockwise-rotation cells must shift their retinal focus-tuning curves downward. (Reprinted, with permission, from Bradley et al. 1996a [copyright AAAS].)

lar focus-tuning curves to the expansion stimuli and compensated the focus tuning in the correct directions (orthogonal to pursuit for rotation and opposite pursuit for contraction). The cross-correlation results on the population data also showed that shifts were similar for all three patterns of motion.

This finding—that pursuit compensation also occurs for rotation- and contraction-sensitive MSTd neurons—has interesting implications for its role in perceptual functions. Fixation of a point straight ahead while moving over a ground plane leads to expansion on the retina. Points tracked farther off to the side lead to increasing amounts of rotation. This is due to the fact that eye-rotation components about an axis orthogonal to the ground plane make the plane appear to rotate. Thus, outward spirals are generally seen (expansion + rotation) when tracking a point on the ground. The focus of such a spiral can in principle tell us the direction of heading, as long as the eye velocity is taken into account. The same arguments pertain to contraction and inward spiral stimuli when looking in a direction opposite to the direction of motion. Thus, the correction of the focus tuning of MSTd neurons for motion patterns other than expansions during eye movements is consistent with their assisting in the mechanism of heading computation.

A second important perceptual function may also be served by the compensation for multiple patterns of motion stimuli. The shifts in focus tuning for all flow stimuli indicate a more general phenomenon of perceptual stability in the face of retinal image motions due to pursuit eye movements. Thus, for instance, when the eyes track across a moving wheel, the wheel does not appear to move up or down. Thus, MSTd may play a general role in compensating for self-induced motion using eye-movement signals. One important outcome of this compensation is the ability to compute heading direction, and another is the perceptual stability of motions in the environment.

Summary

Psychophysical studies have shown that humans use optic flow for heading perception and that heading

judgments remain accurate even during smooth pursuit eye movements. In the absence of depth cues, this computation is only possible if eye velocity is taken into account. Recent physiological studies show that area MSTd, whose neurons combine optic-flow and eye-movement information, is central to heading computation. Expansion-selective neurons in this area are tuned for the retinal position of an expansion focus, and during pursuit, their focus tuning shifts in a way that compensates for the flow added by the eye movement. As a result, these neurons encode the direction of heading both when the eyes are still and when they are moving. Preliminary data suggest that MSTd accounts for head movements as well, in this case by integrating a vestibular signal. Surprisingly, nearly identical results were obtained in neurons selective for rotation and contraction. This suggests that a broad class of optic-flow patterns—not just expansion—may contribute to heading computation.

THREE-DIMENSIONAL-SFM PERCEPTION

Position Versus Velocity Cues

There are two broad theories for SFM perception: one that uses velocity measurements (Clocksin 1980; Longuet-Higgins and Prazdny 1980; Koenderink and Van Doorn 1986; Husain et al. 1989; Treue et al. 1993, 1995; Hildreth et al. 1995) and one that uses position measurements (Ullman 1984; Rieger and Lawton 1985; Grzywacz and Hildreth 1987; Grzywacz et al. 1988; Shariat and Price 1990) to derive three-dimensional structure. Perhaps the best-known position-measurement model is the incremental rigidity algorithm originally formulated by Ullman (1984). This algorithm samples position information derived from a few discrete image views of a moving object and attempts to find a rigid three-dimensional interpretation from the two-dimensional sample frames. Velocity-based algorithms measure the local velocities of points on an image and use the global velocity field to compute three-dimensional SFM. In our laboratory, we explored how humans perceive SFM using a novel stimulus designed to differentiate between position- and velocity-based algorithms. The results of these experiments support the use of velocity and not position measurements (Treue et al. 1993, 1995; Hildreth et al. 1995). Monkeys were also trained to perform similar SFM tasks using this stimulus. When area MT was lesioned, the monkeys could no longer perform this task (Siegel and Andersen 1986). Area MT cells are sensitive to the velocity of stimuli (Maunsell and Van Essen 1983a) and would in general be an unlikely area for making precise position measurements. These experiments establish that the brain uses velocity measurements for the SFM computation. They also establish area MT as a likely site for SFM processing.

Surface Reconstruction

Recent studies from our laboratory suggest that part of the SFM process is the reconstruction of surfaces (Husain et al. 1989; Hildreth et al. 1995; Treue et al. 1995). These studies indicate that the brain forms a mental image of a moving surface, based on velocity measurements. We have modeled this reconstruction process with three stages (Hildreth et al. 1995). The first is the measurement of motion signals, which are made independently of each other. The second step, the segregation of surfaces, is important for interpolation over areas with sparse features while still respecting surface borders. In addition, under transparent conditions or at motion borders, there are often motions in different directions that should not be averaged, but rather assigned to separate surfaces. A third stage then assigns three-dimensional depth values across the surface through the use of motion gradient information. As shown below, neurophysiological studies suggest that macaque areas V1, MT, and MST may carry out functions analogous to the proposed model stages. Specifically, V1 neurons have properties suitable for detecting motion signals, whereas MT neurons probably have a role in surface segmentation and integration. Finally, MT neurons, which are selective for motion gradients (Treue and Andersen 1996), or MST neurons, which are sensitive to motion patterns, may compute depth from motion gradients.

Transparency

Motion transparency—the perception of multiple directions of motion in the same location—is an important part of SFM processing. Transparent stimuli have proven invaluable in exploring the surface reconstruction process and SFM perception. It has long been appreciated that motion thresholds are higher under transparent conditions (Snowden et al. 1991). In physiology experiments, we found that direction-selective V1 neurons generally give the same response to a stimulus moving in their preferred direction, whether or not a second stimulus is present and moving in the opposite direction (Snowden et al. 1991; Qian and Andersen 1994). On the other hand, we found strong suppression for MT neurons under these transparent conditions (Snowden et al. 1991). This result supports the idea that the motion pathway contains two stages, one in V1 that measures motion signals, and a second, opponent stage in MT. We proposed that this opponent stage is part of a surface-reconstruction process and functions to suppress noise such as motion flicker.

The two-stage motion-detection hypothesis suggests that there might exist motion stimuli that would be completely “balanced,” containing motion signals in opposite directions at every local region in the display, which would completely erase the perception of motion. We developed several displays that all eliminated

perceived motion; one of these consisted of two random dot patterns in which the dots moving in opposite directions were all paired with one another. This stimulus produced greater suppression in area MT than in V1, providing strong evidence that area MT is the locus of perception of motion transparency (Qian and Andersen 1994). Furthermore, the degree of suppression of individual MT neurons correlates negatively with their response to visual noise, supporting our proposal that one purpose of the opponency is to suppress noise during surface reconstruction.

Integration of Depth and Stereo

In our psychophysical studies with balanced, non-transparent stimuli, we found that if the two surfaces were separated in depth, the perception of the two surfaces re-emerged (Qian et al. 1994a). Physiological experiments have established that area MT neurons are selective not only to motion, but also to stereoscopic depth (Maunsell and Van Essen 1983b). We reasoned that the improved perception of transparency with stereoscopic depth separation may be a result of the opponent suppression in MT operating primarily within the same stereo planes. Our experiments showed that in fact this is true—inhibition in MT occurs mainly between motion signals with similar disparity (Bradley et al. 1995). This is consistent with opponency operating to minimize random motion signals from a given surface (e.g., due to flicker), whereas stereospecificity prevents opponency between motion signals from different objects at different visual depths. In Figure 6 this basic MT network is shown, with excitatory connections between opponent directions at different stereoscopic depths (D.C. Bradley et al., unpubl.).

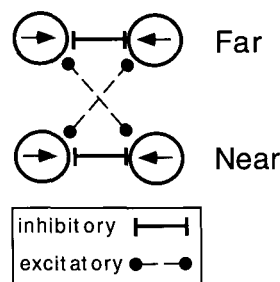


Figure 6. Model of MT motion processing. Neurons tuned to opposite directions inhibit each other within a given disparity channel (*solid lines*) and excite each other between disparity channels (*dashed lines*). Since very few MT neurons are tuned for near-zero disparities, this channel is omitted. However, most near- and far-tuned MT cells respond at least somewhat to zero disparity. What is the predicted response to a stimulus with opposite directions, both at zero disparity? Either direction could be represented in either the near or far channels. However, the preferred configuration would be one in which the two directions occupy *different* disparity channels. Therefore, it is possible that basic depth percepts in SFM stimuli result from depth-specific inhibition and facilitation in MT.

A stereospecific opponency was also proposed in a modeling paper by Nawrot and Blake (1989, 1990) based on psychophysical results showing depth-specific motion adaptation. Another important prediction of the Nawrot and Blake study was that the impression of depth might arise due to these opponent interactions alone (Hiris and Blake 1996). When we see opposite directions at the same disparity, as with two planes or the cylinder stimulus, MT neurons tuned to these opposing motions mutually inhibit each other, possibly driving activity into the near and far channels. This is possible because most MT neurons have broad disparity tuning that extends into the zero-disparity range. The excitatory connections across disparity planes for different motion directions, combined with opponency within stereoplanes, prevent both the near and far cells from representing the same direction (surface). We might thus perceive a depth ordering of the surfaces due to the constraint in MT that different directions must be represented in different disparity channels.

Recent work in our laboratory showed that MT neurons are sensitive to velocity gradients (Treue and Andersen 1996), but the effects are modest. Thus, it is unclear whether MT is also responsible for the depth percept that occurs as a result of motion gradients. Another possible site for this function is MST. Cells in this region, especially MSTd cells, are selective for various types of motion patterns and might be expected to play a role in the third stage of SFM computation; namely, the assignment of three-dimensional shape to surfaces. These MST neurons may be constructed from gradient-selective neurons within MT (Treue and Andersen 1996).

Neural Correlates of SFM Perception in MT

As discussed above, lesions to MT in macaques irreversibly suppress the perception of SFM, suggesting that MT has a major role in SFM computation. However, it could be that MT is needed simply to supply image velocity estimates to another area that computes SFM. To distinguish between these possibilities, we tested for trial-by-trial correlation between MT activities and the percept of SFM in behaving monkeys (Chang et al. 1996). Monkeys were trained to view a revolving cylinder, then make a saccade to indicate the perceived direction of motion of the front surface of the cylinder. Some cylinders contained disparity cues to specify the cylinder's rotation direction, whereas others—the bistable cylinders—contained no disparity and could therefore be seen as turning in either direction. Our preliminary findings suggest that MT responses are indeed correlated with the perceived rotation direction of these bistable cylinders. Therefore, MT's role in SFM perception does not appear limited to velocity estimation, but may in fact be central to the computation of depth from two-dimensional motion signals.

Summary

Humans are able to reconstruct the three-dimensional structure of an object based solely on two-dimensional motion information. This remarkable computation is known as structure-from-motion (SFM). Psychophysical studies have shown that SFM derives from velocity, rather than position, measurements. Moreover, surface interpolation, whereby similar motions are grouped, is a fundamental step in the SFM calculation. Physiological studies suggest that surface movements are computed in a hierarchical fashion, with component motions measured in area V1 and integrated in MT. Recent experiments on the mechanism of this integration have revealed a property of MT cells that may be essential to the SFM percept; namely, opposite directions are mutually inhibitory *within a given disparity channel*. As a result, MT does not readily represent multiple directions at the same depth, and MT neurons may thus be forced to encode opposing directions at different depths. Therefore, it is possible that the basic inference of depth in SFM stimuli stems directly from the stereospecificity of opponent interactions in MT. It is not clear whether MT computes depth veridically in these stimuli using motion gradients. There is evidence for gradient selectivity in MT neurons, but the effect is moderate compared to the cell's selectivity for direction of disparity. It is also possible that MST, which contains neurons sensitive to motion patterns, contributes to perceived depth from motion gradients.

CONCLUSION

Studies on heading and SFM perception have taught us a great deal about how different kinds of information converge on the primate motion pathway, allowing us to extract useful information from the visual world. One important observation is that even early stages of cortical processing are involved in rather specific computations. MT, for example, appears to have a central role in SFM computation, and MST is very likely involved in heading perception. Both of these computations are elaborate and highly relevant to our survival in a changing environment. This is not to say that cortical areas are specialized for a single computation. Indeed, it is probable that MT participates in heading analysis by providing velocity estimates to MSTd, whereas MST itself may help compute SFM. However, as experiments have gone beyond simple, unidirectional stimuli to more realistic displays that provide multiple retinal and extraretinal cues, it has become clear that the cortical motion areas combine these cues in ways that suggest a particular type of computation. Therefore, to understand the function of cortical areas, it is essential that they be studied in the context of a specific computational goal.

ACKNOWLEDGMENTS

We thank Sylvie Gertmenian, Betty Gillikin, and Jason Liao for their excellent technical assistance. This work was supported by the National Eye Institute, the Sloan Center for Theoretical Neurobiology at Caltech, the Office of Naval Research, and the Human Frontiers Scientific Program.

REFERENCES

- Bradley, D.C., N. Qian, and R.A. Andersen. 1995. Integration of motion and stereopsis in middle temporal cortical area of macaques. *Nature* **373**: 609.
- Bradley, D.C., M. Maxwell, R.A. Andersen, M.S. Banks, and K.V. Shenoy. 1996a. Neural mechanisms for heading perception in primate visual cortex. *Science* **273**: 1544.
- . 1996b. Heading computation during pursuit eye movements in cortical area MSTd. *Soc. Neurosci. Abstr.* **22**: (in press).
- Chang, G.C., D.C. Bradley, and R.A. Andersen. 1996. Neural correlate of 3-D structure from motion (SFM) perception in area MT. *Soc. Neurosci. Abstr.* **22**: (in press).
- Cocks, W.F. 1980. Perception of surface slant and edge labels from optical flow: A computational approach. *Perception* **9**: 253.
- Duffy, C.J. and R.H. Wurtz. 1991a. Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *J. Neurophysiol.* **65**: 1329.
- . 1991b. Sensitivity of MST neurons to optic flow stimuli. II. Mechanisms of response selectivity revealed by small-field stimuli. *J. Neurophysiol.* **65**: 1346.
- . 1995. Response of monkey MST neurons to optic flow stimuli with shifted centers motion. *J. Neurosci.* **15**: 5192.
- Geesaman, B.J. and R.A. Andersen. 1996. The analysis of complex motion patterns by form/cue invariant MSTd neurons. *J. Neurosci.* **16**: 4716.
- Gibson, J.J. 1950. *The perception of the visual world*. Houghton Mifflin, Boston, Massachusetts.
- Graziano, M.S.A., R.A. Andersen, and R.J. Snowden. 1994. Tuning of MST neurons to spiral motions. *J. Neurosci.* **14**: 54.
- Grzywacz, N.M. and E.C. Hildreth. 1987. Incremental rigidity scheme for recovering structure from motion-position-based versus velocity-based formulations. *J. Opt. Soc. Am.* **4**: 503.
- Grzywacz, N.M., E.C. Hildreth, V.K. Inada, and E.H. Adelson. 1988. The temporal integration of 3-D structure from motion: A computational and psychophysical study. In *Organization of neural networks* (ed. W. Von Seelen et al.) p. 239. VCH, Weinheim, Germany.
- Hatsopoulos, N.G. and W.H. Warren. 1991. Visual navigation with a neural network. *Neural Networks* **4**: 303.
- Hildreth, E.C. 1992. Recovering heading for visually-guided navigation. *Vision Res.* **32**: 1177.
- Hildreth, E.C., H. Ando, R.A. Andersen, and S. Treue. 1995. Recovering three-dimensional structure from motion with surface reconstruction. *Vision Res.* **35**: 117.
- Hiris, E. and R. Blake. 1996. Direction repulsion in motion transparency. *Visual Neurosci.* **13**: 187.
- Husain, M., S. Treue, R.A. Andersen. 1989. Surface interpolation in three-dimensional structure-from-motion perception. *Neural Comp.* **1**: 324.
- Kawano, K. and M. Sasaki. 1984. Response properties of neurons in posterior parietal cortex of monkey during visual-vestibular stimulation. II. Optokinetic neurons. *J. Neurophysiol.* **51**: 352.
- Kawano, K., M. Sasaki, and M. Yamashita. 1980. Vestibular input to visual tracking neurons in the posterior parietal association cortex of the monkey. *Neurosci. Lett.* **17**: 55.
- . 1984. Response properties of neurons in posterior parietal cortex of monkey during visual-vestibular stimulation. I. Visual tracking neurons. *J. Neurophysiol.* **51**: 340.
- Koenderink, J.J. and A.J. Van Doorn. 1981. Exterspecific component of the motion parallax field. *J. Opt. Soc. Am.* **71**: 953.
- . 1986. Depth and shape from differential perspective in the presence of bending deformations. *J. Opt. Soc. of Am.* **A3**: 242.
- Komatsu, H. and R.H. Wurtz. 1988a. Relation of cortical areas MT and MST to pursuit eye-movements. I. Localization and visual properties of neurons. *J. Neurophysiol.* **60**: 580.
- . 1988b. Relation of cortical areas MT and MST to pursuit eye-movements. III. Interaction with full-field visual stimulation. *J. Neurophysiol.* **60**: 621.
- Lagae, L., H. Maes, S. Raiguel, D.K. Xiao, and G.A. Orban. 1994. Responses of macaque STS neurons to optic flow components—A comparison of areas MT and MST. *J. Neurophysiol.* **71**: 1597.
- Lappe, M. and J.P. Rauschecker. 1994. Heading detection from optic flow. *Nature* **369**: 712.
- Longuet-Higgins, H.C. and K. Prazdny. 1980. The interpretation of a moving retinal image. *Proc. R. Soc. Lond. B Biol. Sci.* **211**: 151.
- Lynch, J.C., V.B. Mountcastle, W.H. Talbot, and T.C.T. Yin. 1977. Parietal lobe mechanisms for directed visual attention. *J. Neurophysiol.* **40**: 362.
- Maunsell, J. and D.C. Van Essen. 1983a. Functional properties of neurons in middle temporal visual area of the macaque monkey. I. Selectivity for stimulus direction, speed, and orientation. *J. Neurophysiol.* **49**: 1127.
- . 1983b. Functional-properties of neurons in middle temporal visual area of the macaque monkey. II. Binocular interactions and sensitivity to binocular disparity. *J. Neurophysiol.* **49**: 1148.
- Mountcastle, V.B., J.C. Lynch, A. Georgopoulos, H. Sakata, and C. Acuna. 1975. The posterior parietal association cortex of the monkey: Command functions for operations within extrapersonal space. *J. Neurophysiol.* **38**: 871.
- Nawrot, M. and R. Blake. 1989. Neural integration of information specifying structure from stereopsis and motion. *Science* **244**: 716.
- . 1990. A neural network model of kinetic depth. *Visual Neurosci.* **6**: 219.
- Newsome, W.T., R.H. Wurtz, and H. Komatsu. 1988. Relation of cortical areas MT and MST to pursuit eye-movements. II. Differentiation of retinal from extraretinal inputs. *J. Neurophysiol.* **60**: 604.
- Perrone, J.A. and L.S. Stone. 1994. A model of self-motion estimation within primate extrastriate visual-cortex. *Vision Res.* **34**: 2917.
- Qian, N. and R.A. Andersen. 1994. Transparent motion perception as unbalanced signals: II. Physiology. *J. Neurosci.* **14**: 7367.
- Qian, N., R.A. Andersen, and E.H. Adelson. 1994a. Transparent motion perception as detection of unbalanced motion signals: I. Psychophysics. *J. Neurosci.* **14**: 7357.

- . 1994b. Transparent motion perception as detection of unbalanced motion signals: III. Modeling. *J. Neurosci.* **14**: 7381.
- Regan, D. and K.I. Beverley. 1982. How do we avoid confounding the direction we are looking and the direction we are moving. *Science* **215**: 194.
- Rieger, J.H. and D.T. Lawton. 1985. Processing differential image motion. *J. Opt. Soc. Am.* **A2**: 354.
- Royden, C.S., M.K.S. Banks, and J.A. Crowell. 1992. The perception of heading during eye movements. *Nature* **360**: 583.
- Royden, C.S., J.A. Crowell, and M.K.S. Banks. 1994. Estimating heading during eye movements. *Vision Res.* **34**: 3197.
- Saito, H., M. Yukie, K. Tanaka, K. Hikosaka, Y. Fukada, and E. Iwai. 1986. Integration of direction signals of image motion in the superior temporal sulcus of the macaque monkey. *J. Neurosci.* **6**: 145.
- Sakata, H., H. Shibutani, K. Kawano, and T. Harrington. 1985. Neural mechanisms of space vision in the parietal association cortex of the monkey. *Vision Res.* **25**: 453.
- Shariat, H. and K.E. Price. 1990. Motion estimation with more than 2 frames. *IEEE Trans. Patt. Anal. Mach. Intell.* **12**: 417.
- Shenoy, K.V., D.C. Bradley, and R.A. Andersen. 1996. Heading computation during head movements in macaque cortical area MSTd. *Soc. Neurosci. Abstr.* **22**: (in press).
- Siegel, R.M. and R.A. Andersen. 1986. Motion perceptual deficits following ibotenic acid lesions of the middle temporal area in behaving rhesus monkey. *Soc. Neurosci. Abstr.* **12**: 1183.
- Snowden, R.J., S. Treue, R.G. Erickson, and R.A. Andersen. 1991. The response of area MT and V1 neurons to transparent motion. *J. Neurosci.* **11**: 2768.
- Stone, L.S. and J.A. Perrone. 1996. Translation and rotation trade-off in human visual heading estimation. *Invest. Ophthalmol.* **37**: 2359. (Abstr.).
- Tanaka, K. and H.A. Saito. 1989. Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. *J. Neurophysiol.* **62**: 626.
- Tanaka, K., Y. Fukada, and H.A. Saito. 1989. Underlying mechanisms of the response specificity of expansion/contraction and rotation cells in the dorsal part of the medial superior temporal area of the macaque monkey. *J. Neurophysiol.* **62**: 642.
- Tanaka, K., K. Hikosaka, H. Saito, M. Yukie, Y. Fukada, and E. Iwai. 1986. Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey. *J. Neurosci.* **6**: 134.
- Thier, P. and R.C. Ericksen. 1992. Responses of visual tracking neurons from cortical area MST-1 to visual, eye and head motion. *Eur. J. Neurosci.* **4**: 539.
- Treue, S. and R.A. Andersen. 1996. Neural responses to velocity gradients in macaque cortical area MT. *Visual Neurosci.* **13**: 797.
- Treue, S., R. Snowden, and R.A. Andersen. 1993. The effect of transiency on perceived velocity of visual patterns: A case of "temporal capture." *Vision Res.* **33**: 791.
- Treue, S., R.A. Andersen, H. Ando, and E.C. Hildreth. 1995. Structure-from-motion: Perceptual evidence for surface interpolation. *Vision Res.* **35**: 139.
- Ullman, S. 1984. Maximizing rigidity: The incremental recovery of 3-D structure from rigid and nonrigid motion. *Perception* **13**: 255.
- Warren, W.H. 1995. Self-motion: Visual perception and visual control. In *Perception of space and motion* (ed. W. Epstein and S.J. Rogers), p. 263. Academic Press, San Diego.
- Warren, W.H. and D.J. Hannon. 1988. Direction of self-motion is perceived from optical flow. *Nature* **336**: 162.